CLAIMS

What is claimed is:

1   1.   A computer-implemented method for identifying optimal allocations of computing
2   resources in a data processing arrangement having a plurality of computing machines that
3   host a plurality of application processes, comprising:
4   establishing a plurality of server models, each server model including one or more
5   server nodes, wherein each server node has an associated set of capacity attributes;
6   designating a layered relationship between the server models, wherein for a first
7   server-model layer immediately above a second server-model layer, the second server-
8   model layer includes respective models that represent the nodes in the first server-model
9   layer;
10   establishing a plurality of service models, each service model including one or
11   more service nodes, wherein each service node has an associated set of demand attributes;
12   designating a layered relationship between the service models, wherein for a first
13   service-model layer immediately above a second service-model layer, the second service-
14   model layer includes respective models that represent the nodes in the first server-model
15   layer;
16   normalizing the capacity attributes and the demand attributes; and
17   generating an optimized mapping of service nodes in a user-selected service model
18   to server nodes in a user-selected server model as a function of the associated sets of
19   demand and capacity attributes.

1   2.   The method of claim 1, wherein the capacity attributes include processing capacity
2   and storage capacity and the demand attributes include processing demand and storage
3   demand.

1   3.   The method of claim 2, further comprising:
2   establishing one or more service-node relationships between selected pairs of the
3   service nodes, wherein each service-node relationship has an associated transport demand
4   attribute specifying a quantity of communication resources required for communication
5   between the associated pair of service nodes;

20

6   establishing one or more server-node relationships between selected pairs of the

7   server nodes, wherein each server-node relationship has an associated transport capacity

8   attribute specifying a quantity of communication resources available for communication

9   between the associated pair of server nodes;

10   normalizing the transport demand attributes; and

11   generating the optimized mapping as a function of the service-node relationships

12   and server-node relationships.

1   4.   The method of claim 3, further comprising:

2   selecting one of the server nodes as a base server node;

3   normalizing the capacity attributes relative to the capacity attribute of the base

4   server node;

5   selecting one of the service nodes as a base service node; and

6   normalizing the demand attributes relative to the demand attribute of the base

7   service node.

1   5.   The method of claim 4, wherein each service node has an associated set of capacity

2   attributes and further comprising generating an optimized mapping of service nodes in a

3   first user-selected service model to service nodes in a second user-selected service model

4   as a function of the demand attributes of the first service model and capacity attributes of

5   the second service model.

1   6.   The method of claim 5, wherein each server node has an associated set of demand

2   attributes and further comprising generating an optimized mapping of server nodes in a

3   first user-selected server model to server nodes in a second user-selected server model as a

4   function of the demand attributes of the first server model and capacity attributes of the

5   second server model.

1   7.   The method of claim 1, further comprising:

2   establishing one or more service-node relationships between selected pairs of the

3   service nodes, wherein each service-node relationship has an associated transport demand

4   attribute specifying a quantity of communication resources required for communication

5   between the associated pair of service nodes;

6   establishing one or more server-node relationships between selected pairs of the

7   server nodes, wherein each server-node relationship has an associated transport capacity

8   attribute specifying a quantity of communication resources available for communication

9   between the associated pair of server nodes;

10   normalizing the transport demand attributes; and

11   generating the optimized mapping as a function of the service-node relationships

12   and server-node relationships.

1   8.   The method of claim 1, further comprising:

2   selecting one of the server nodes as a base server node;

3   normalizing the capacity attributes relative to corresponding capacity attributes of

4   the base server node;

5   selecting one of the service nodes as a base service node; and

6   normalizing the demand attributes relative to corresponding demand attributes of

7   the base service node.

1   9.   The method of claim 1, wherein each service node has an associated set of capacity

2   attributes and further comprising generating an optimized mapping of service nodes in a

3   first user-selected service model to service nodes in a second user-selected service model

4   as a function of the demand attributes of the first service model and capacity attributes of

5   the second service model.

1   10.   The method of claim 1, wherein each server node has an associated set of demand

2   attributes and further comprising generating an optimized mapping of server nodes in a

3   first user-selected server model to server nodes in a second user-selected server model as a

4   function of the demand attributes of the first server model and capacity attributes of the

5   second server model.

1   11.   An apparatus for identifying optimal allocations of computing resources in a data

2   processing arrangement having a plurality of computing machines that host a plurality of

3   application processes, comprising:

4   means for establishing a plurality of server models, each server model including

5   one or more server nodes, wherein each server node has an associated set of capacity

6   attributes;

7   means for designating a layered relationship between the server models, wherein

8   for a first server-model layer immediately above a second server-model layer, the second

9   server-model layer includes respective models that represent the nodes in the first server-

10   model layer;

11   means for establishing a plurality of service models, each service model including

12   one or more service nodes, wherein each service node has an associated set of demand

13   attributes;

14   means for designating a layered relationship between the service models, wherein

15   for a first service-model layer immediately above a second service-model layer, the second

16   service-model layer includes respective models that represent the nodes in the first server-

17   model layer;

18   means for normalizing the capacity attributes and the demand attributes; and

19   means for generating an optimized mapping of service nodes in a user-selected

20   service model to server nodes in a user-selected server model as a function of the

21   associated sets of demand and capacity attributes.

1   12.   A system for identifying optimal allocations of computing resources in a data

2   processing arrangement having a plurality of computing machines that host a plurality of

3   application processes, comprising:

4   a model repository including a plurality of server models and a plurality of service

5   models, each server model including one or more server nodes and each server node

6   having an associated set of normalized capacity attributes, each service model including

7   one or more service nodes and each service node having an associated set of normalized

8   demand attributes, wherein the server models are defined in a layered relationship and for

9   a first server-model layer immediately above a second server-model layer, the second

10   server-model layer includes respective models that represent the nodes in the first server-

11   model layer, and the service models are defined in a layered relationship and for a first

12   service-model layer immediately above a second service-model layer, the second service-

13   model layer includes respective models that represent the nodes in the first service-model

14   layer; and

15   an optimization engine coupled to the model repository, the optimization engine

16   configured to generate an optimized mapping of service nodes in a user-selected service

17   model to server nodes in a user-selected server model as a function of the associated

18   normalized demand and capacity attributes.

1   13.   The system of claim 12, further comprising:

2   wherein the model repository further includes one or more service-node

3   relationships between selected pairs of the service nodes, each service-node relationship

4   having an associated transport demand attribute that specifies a normalized quantity of

5   communication resources required for communication between the associated pair of

6   service nodes;

7   wherein the model repository further includes one or more server-node

8   relationships between selected pairs of the server nodes, each server-node relationship

9   having an associated transport capacity attribute that specifies a normalized quantity of

10   communication resources available for communication between the associated pair of

11   server nodes; and

12   the optimization engine is further configured to generate the optimized mapping as

13   a function of the service-node relationships and server-node relationships.

1   14.   The system of claim 12, wherein each service node has an associated set of

2   normalized capacity attributes and the optimization engine is further configured to

3   generate an optimized mapping of service nodes in a first user-selected service model to

4   service nodes in a second user-selected service model as a function of the demand

5   attributes of the first service model and capacity attributes of the second service model.

10010928-1

1   15.   The system of claim 12, wherein each server node has an associated set of

2   normalized demand attributes and the optimization engine is further configured to generate

3   an optimized mapping of server nodes in a first user-selected server model to server nodes

4   in a second user-selected server model as a function of the demand attributes of the first

5   server model and capacity attributes of the second server model.